Cogn-IQ Research Papers Research Article ◆ Open Access doi: pubscience.org/ps-1n0I3-799113-Lpri

CONFIRMATORY FACTOR ANALYSIS OF THE JCCES

Confirmatory Factor Analysis of the Jouve Cerebrals Crystallized Educational Scale: Evidence for a Three-Factor Structure With a Dominant General Factor

Xavier Jouve
Head of Research
Cogn-IQ

Author Note

Xavier Jouve, Head of Research, Cogn-IQ https://orcid.org/0009-0006-1283-045X

The author is the developer of the Jouve Cerebrals Crystallized Educational Scale (JCCES) analyzed in this article. This potential conflict of interest is disclosed in the interest of full transparency. No external funding was received for this work.

Correspondence concerning this article should be addressed to Xavier Jouve, Cogn-IQ. Email: research@cogn-iq.org

ABSTRACT

This study examined the latent factor structure of the Jouve Cerebrals Crystallized Educational Scale (JCCES) using confirmatory factor analysis (CFA) with a sample of 1,553 examinees who completed 129 items organized into three subtests: Verbal Analogies (VA, 41 items), Mathematical Problems (MP, 32 items), and General Knowledge (GK, 56 items). Item parcels were constructed to create 11 continuous indicators (4 VA, 3 MP, 4 GK), and four competing structural models were evaluated: (a) a unidimensional model, (b) a three-factor correlated model, (c) a higher-order *g* model, and (d) a bifactor model. The three-factor correlated model ($\chi^2 = 1{,}010.16$, $df = 41$, CFI = .911, RMSEA = .123) fit significantly better than the one-factor model ($\Delta\chi^2 = 1{,}370$, $\Delta df = 3$, $p < .001$). Interfactor correlations were substantial (VA–MP = .694, VA–GK = .890, MP–GK = .609), indicating a strong common core. Schmid–Leiman decomposition of the three-factor model yielded omega hierarchical (ω_h) = .879, omega total (ω_t) = .937, and an explained common variance (ECV) of .826; the general factor therefore accounts for 87.9% of total-score variance and 93.7% of its reliable variance. The Mathematical Problems subtest retained the most unique reliable variance ($\omega_s = .260$), while Verbal Analogies ($\omega_s = .069$) and General Knowledge ($\omega_s = .128$) were largely subsumed by the general factor. Because conventional fit cut-offs are miscalibrated for parcelled binary indicators, a benchmark

was simulated from a model correct by construction: it attains RMSEA = .063 and SRMR = .025 in this design, against the observed .123 and .056, indicating that the elevated RMSEA is part artefact and part genuine misspecification, localized in the covariance among the easiest parcels. Practical implications for test interpretation are discussed.

Keywords: confirmatory factor analysis, structural equation modeling, JCCES, crystallized intelligence, factor structure

Introduction

The Jouve Cerebrals Crystallized Educational Scale (JCCES; Jouve et al., 2023) is a 129-item measure of crystallized intelligence (*Gc*) comprising three subtests: Verbal Analogies (VA), Mathematical Problems (MP), and General Knowledge (GK). Previous psychometric evaluations have established high internal consistency for the full scale ($\alpha = .961$) and its subtests (VA $\alpha = .853$, MP $\alpha = .914$, GK $\alpha = .942$), and a two-parameter logistic (2PL) item response theory analysis confirmed adequate item discrimination across the ability continuum (Jouve, 2023). However, the dimensionality of the JCCES has not been examined using confirmatory methods. The present study addresses this gap by fitting competing structural models to item parcel data from a large sample of examinees. Because the question is not only whether the domains separate but how much of the reliable variance a general factor claims, the analysis is framed throughout in the model-based reliability indices — ω_h, ω_s, and the explained common variance — that a bifactor decomposition supplies (Reise, 2012; Rodriguez et al., 2016), rather than in fit indices alone.

The construct of crystallized intelligence, as defined within the Cattell–Horn–Carroll (CHC) framework, encompasses knowledge and skills acquired through education and acculturation (Cattell, 1963; McGrew, 2009). CHC theory distinguishes *Gc* from fluid intelligence (*Gf*), but within the crystallized domain, the degree to which verbal, quantitative, and knowledge abilities form separable factors versus a single undifferentiated construct remains an empirical question. The JCCES, by design, samples three distinct content domains—verbal reasoning, mathematical problem

solving, and general factual knowledge—making it a suitable instrument for investigating this question.

This study provides the first confirmatory factor-analytic examination of the JCCES latent structure. We ask:

RQ1. Is the JCCES better described by a unidimensional model or by a three-factor structure corresponding to its Verbal Analogies, Mathematical Problems, and General Knowledge subtests?
RQ2. How much of the reliable variance is attributable to a general factor versus the specific subtest factors?

The CHC Framework and the Dimensionality of Crystallized Intelligence

Within the Cattell–Horn–Carroll (CHC) taxonomy, crystallized intelligence (*Gc*) is conceptualized as a broad stratum-II ability reflecting the depth and breadth of knowledge and skills acquired through education and acculturation (Carroll, 1993; Cattell, 1963; McGrew, 2009). *Gc* encompasses several narrow abilities, including lexical knowledge, general verbal information, language development, and communication ability. Quantitative reasoning (*Gq*) occupies a somewhat ambiguous position in the CHC hierarchy: some formulations treat *Gq* as a facet of *Gc* (reflecting educationally acquired mathematical knowledge), while others position it as a partially distinct stratum-II ability that shares variance with both *Gc* and fluid intelligence (*Gf*; McGrew, 2009). This theoretical ambiguity makes the JCCES—which measures verbal, quantitative, and knowledge content within a single battery—an informative instrument for investigating the internal structure of the crystallized domain.

The question of whether broad cognitive abilities decompose into separable narrow factors or function as essentially unidimensional constructs has been extensively debated. Carroll's (1993) three-stratum model posits that individual differences in cognitive performance reflect both a general factor (*g*) and group factors that account for ability-specific variance. Johnson et al. (2004) demonstrated that *g* extracted from diverse

test batteries is highly stable across different sets of indicators, supporting the robust reality of the general factor. At the same time, group factors such as *Gc* and *Gq* consistently emerge as separable sources of variance in large-scale factor analytic studies. The practical implication is that a well-constructed crystallized intelligence battery should yield subtest scores that are substantially correlated (due to *g*) yet retain measurable unique variance corresponding to their distinct content domains.

Confirmatory Factor Analysis and the LISREL Tradition

Confirmatory factor analysis (CFA) within the structural equation modeling (SEM) framework provides a rigorous method for evaluating hypothesized factor structures against observed data (Bollen, 1989; Jöreskog, 1969). Unlike exploratory factor analysis, CFA allows the researcher to specify the number of latent factors, the pattern of factor loadings (which indicators load on which factors), and interfactor relationships, then evaluate model fit using multiple indices. This approach, originating with Jöreskog's LISREL program, has become the standard method for investigating test dimensionality in psychometric research (Kline, 2016).

When observed indicators are binary (correct/incorrect) rather than continuous, direct maximum likelihood (ML) estimation on item-level data can produce biased estimates because the normality assumption is violated. Item parceling—the practice of summing small groups of items into composite indicators—provides a practical solution by creating approximately continuous variables with improved distributional properties (Little et al., 2002; Matsunaga, 2008). Parceling is particularly appropriate when the goal is to evaluate the higher-order factor structure among content domains rather than item-level dimensionality. Bandalos (2002) demonstrated that parceling binary items into continuous composites can substantially reduce parameter estimate bias and improve model convergence under ML estimation, though it also introduces assumptions about within-parcel unidimensionality that must be considered.

The Present Study

The present study evaluated four competing models for the JCCES: (a) a *unidimensional model* positing that all items measure a single latent factor; (b) a *three-factor correlated model* positing separate but correlated VA, MP, and GK factors; (c) a *higher-order model* with a second-order general factor explaining the correlations among the three first-order factors; and (d) a *bifactor model* with a general factor and three orthogonal specific factors. Additionally, omega hierarchical (ω_h) and the explained common variance (ECV) were computed from the Schmid–Leiman decomposition to quantify the relative contributions of the general and specific factors to total score variance (Rodriguez et al., 2016; Zinbarg et al., 2005). Figure 1 gives their path diagrams.

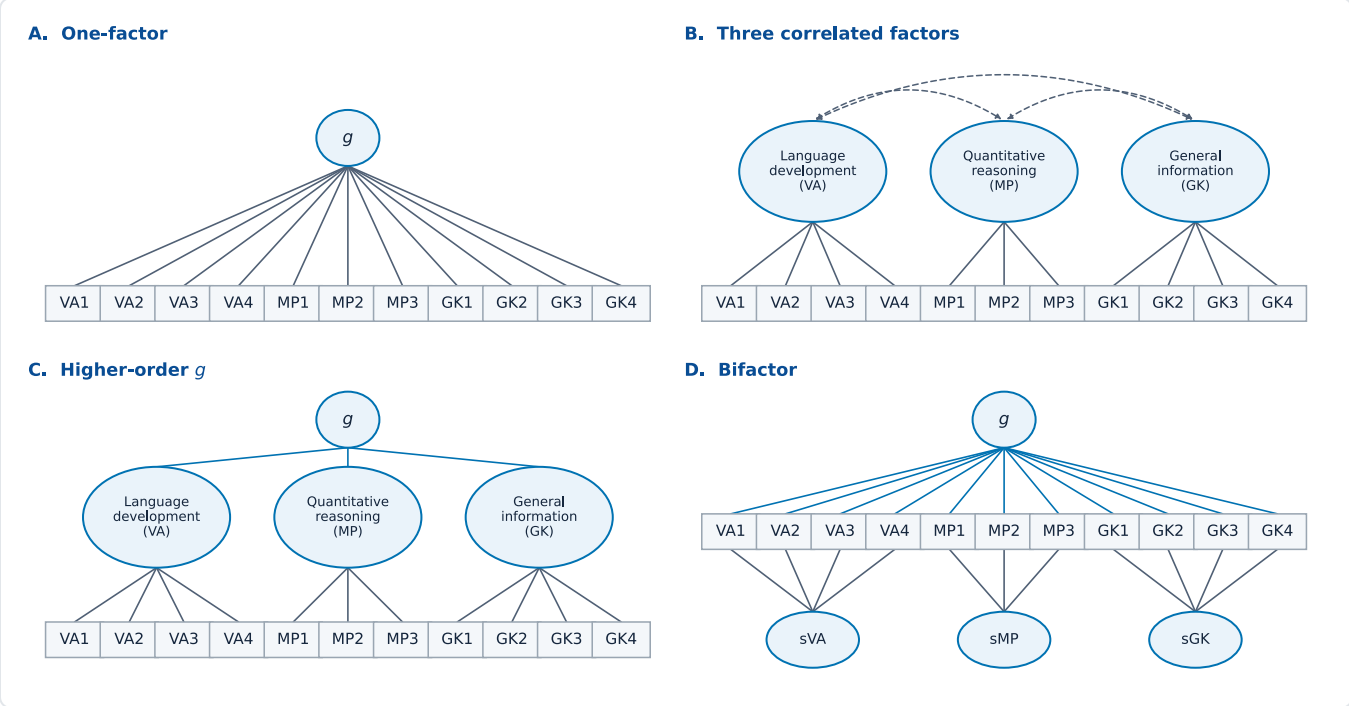

Figure 1. The four competing measurement models, as specified. Rectangles are the 11 item parcels (VA = Verbal Analogies, 4 parcels; MP = Mathematical Problems, 3; GK = General Knowledge, 4). Ellipses are latent factors. Each is named for the CHC narrow ability the subtest indexes rather than for the subtest itself, which the parcels already carry; the subtest abbreviation used in Tables 3–5 follows in parentheses, so the factors in the diagrams and the factors in the tables are plainly the same objects. Dashed double arrows are freely estimated factor correlations; sVA, sMP and sGK are the orthogonal specific factors of the bifactor model. Models B and C differ in claim but not in testable content: with three first-order factors the second-order layer of C is just-identified, so it reproduces exactly the three correlations that B estimates freely.

Based on the CHC theoretical framework and the content composition of the JCCES, we hypothesized that: (a) the three-factor model would fit significantly better than the unidimensional model, reflecting the distinct content domains; (b) interfactor

correlations would be high, consistent with a dominant general factor underlying crystallized performance; and (c) the VA–GK correlation would exceed the VA–MP and MP–GK correlations, reflecting the verbal–educational commonality of verbal analogies and general knowledge tasks relative to mathematical reasoning.

Method

Participants

The sample consisted of 1,553 examinees who completed the JCCES through the Cogn-IQ.org testing platform. All examinees completed the full 129-item test battery under standardized online administration conditions. No responses were missing, and the reason is structural rather than fortunate: the platform submits a value for every item, and an item left blank is scored as incorrect under the operational scoring rule, exactly as it would be in a proctored power test. The analytic matrix is therefore complete by construction, at the cost that omissions are treated as errors rather than as missing data — a conservative choice for a power test in which examinees are instructed to attempt every item. Examinees were adults who voluntarily sought cognitive assessment.

Instrument

The JCCES is a power test of crystallized intelligence comprising 129 items across three subtests. Verbal Analogies (VA; 41 items) presents word pairs and requires the examinee to identify the analogical relationship. Mathematical Problems (MP; 32 items) presents quantitative reasoning problems of increasing difficulty. General Knowledge (GK; 56 items) assesses breadth and depth of factual knowledge across academic domains. Items within each subtest are arranged in approximate order of difficulty. VA and GK subtests use an open-ended response format with automated fuzzy-matching scoring; MP uses constructed-response numerical answers. All items are scored dichotomously (correct/incorrect). In the narrow-ability stratum of the CHC taxonomy (Carroll, 1993; McGrew, 2009), the three subtests correspond most closely to language development (LD) for VA, quantitative reasoning (RQ) for MP, and general verbal information (K0) for GK. The latent factors are labelled by these narrow abilities throughout, since the subtest

names are already carried by the parcels themselves; the correspondence is a classification of item content, not an empirical result of this study.

Item Parceling

Because the 129 items are binary, direct ML estimation at the item level would violate the multivariate normality assumption. Items within each subtest were therefore grouped into parcels by summing blocks of consecutive items, producing approximately continuous indicators suitable for ML (Little et al., 2002). Parcelling suits the question being asked, which is a *domain*-level one: whether Verbal Analogies, Mathematical Problems, and General Knowledge are separable latent factors, and how much of their reliable variance a general factor claims. It is not the question of whether the items within a domain are themselves unidimensional, which this design cannot address and does not claim to. Two features of the scheme bound what it can support. Parcels were formed *within* subtests only, never across them, so the parcelling cannot manufacture the three-factor structure the models are meant to test: cross-domain covariance must still be estimated between the factors. And because each subtest's items are ordered by ascending difficulty, consecutive blocks yield parcels that differ systematically in mean difficulty rather than parcels balanced on it. That was a deliberate simplification with a cost: parcels at the extremes of the difficulty range have restricted variance, which can attenuate their loadings, and the residual correlations reported below show the consequence. The parcels are, per subtest:

- VA: 4 parcels (VA_P1 through VA_P4), comprising 10, 10, 11, and 10 items respectively.
- MP: 3 parcels (MP_P1 through MP_P3), comprising 10, 11, and 11 items respectively.
- GK: 4 parcels (GK_P1 through GK_P4), each comprising 14 items.

This yielded 11 observed indicators for use in the CFA models.

Analytic Strategy

Four CFA models were estimated using maximum likelihood in *semopy* (Igolkina & Meshcheryakov, 2020), a Python structural equation modeling package. Model fit was

evaluated using the chi-square test statistic (χ^2), the comparative fit index (CFI), the Tucker–Lewis index (TLI), and the root mean square error of approximation (RMSEA). Following conventional guidelines (Hu & Bentler, 1999), good fit is indicated by CFI ≥ .95, TLI ≥ .95, and RMSEA ≤ .06; acceptable fit by CFI ≥ .90, TLI ≥ .90, and RMSEA ≤ .08. Nested models were compared using the chi-square difference test, supplemented by the changes in CFI and RMSEA, since with $N = 1{,}553$ the chi-square difference is significant for differences of little practical size; a ΔCFI exceeding .01 is conventionally read as a non-trivial improvement (Cheung & Rensvold, 2002). One estimator decision should be stated at the outset rather than left to the Limitations. Binary items are properly analysed by least squares on tetrachoric correlations rather than by ML, and the estimator question cannot be waved away. ML on parcel sums is retained as the primary analysis because the question is domain-level and the parcel sums — each a count over ten to fourteen items — are close enough to continuous for ML to be defensible on them. But the parcel-level ML analysis is not left to stand alone: the same four models were refitted at the item level to the tetrachoric correlation matrix of all 129 binary items, by unweighted least squares (ULS), and the results are reported alongside. ULS is a consistent estimator for polychoric-based factor models and performs at least as well as diagonally weighted least squares across a wide range of conditions (Forero et al., 2009). One constraint of the toolchain is real and is stated rather than finessed: the mean- and variance-adjusted test statistic that gives WLSMV its second M requires the asymptotic covariance matrix of the 8,256 tetrachoric correlations, which was not computable here. No chi-square-based index is therefore reported for the item-level fit; SRMR, which is a residual summary and carries no distributional assumption, is reported instead and is directly comparable across models. The parcel models yield elevated RMSEA values, and rather than attribute them to the indicators by assertion, the Results calibrate how much of the elevation the indicators can actually account for, by simulating data from a model that is correct by construction and refitting it through the same pipeline. The Schmid–Leiman orthogonalization (Schmid & Leiman, 1957) was applied to the three-factor correlated solution to derive general and specific factor loadings, from which omega hierarchical (ω_h), omega subscale (ω_s), and ECV were computed.

Results

Descriptive Statistics

Table 1 presents descriptive statistics for the three JCCES subtests and the total score. The total score distribution was approximately symmetric (skewness = −0.02, kurtosis = −0.58), as were the subtest score distributions.

Table 1

Descriptive Statistics for JCCES Subtests and Total Score (N = 1,553)

Scale	*k*	*M*	*SD*	Min	Max	Skew.	Kurt.	α
VA	41	23.71	6.21	1	41	−0.23	−0.16	.853
MP	32	13.51	7.07	0	32	−0.02	−0.90	.914
GK	56	27.74	12.12	0	56	0.05	−0.79	.942
Total (CAI)	129	64.96	22.39	5	124	−0.02	−0.58	.961

Note. *k* = number of items; Skew. = skewness; Kurt. = kurtosis; α = Cronbach's alpha. VA = Verbal Analogies; MP = Mathematical Problems; GK = General Knowledge; CAI = Cognitive Acumen Index (JCCES composite).

Subtest intercorrelations were substantial: VA–MP $r = .599$, VA–GK $r = .774$, MP–GK $r = .566$. The high correlation between VA and GK is consistent with both subtests measuring verbal–educational aspects of *Gc*, while MP's somewhat lower correlations with the other subtests are consistent with quantitative reasoning tapping partially distinct abilities.

Eigenvalue Analysis of Item Parcels

An eigenvalue decomposition of the 11×11 parcel correlation matrix revealed three eigenvalues greater than 1.0 (6.01, 1.17, 1.03), accounting for 54.6%, 10.6%, and 9.3% of the total variance, respectively. The dominant first eigenvalue (ratio of first to second: 5.14) suggested a strong general factor, while the presence of three eigenvalues above

unity was consistent with the hypothesized three-factor structure. Cumulatively, the first three components accounted for 74.6% of total variance.

Model Fit Comparison

Table 2 presents fit statistics for the four CFA models, and Figure 2 plots them against the conventional cut-offs.

Table 2
Fit Statistics for Competing CFA Models (N = 1,553)

Model	χ^2	*df*	CFI	TLI	RMSEA	SRMR	Interpretation
1. One-factor	2,380.15	44	.786	.733	.185 [.179, .191]	.086	Rejected; one dimension cannot reproduce the parcel covariances
2. Three-factor correlated	1,010.16	41	.911	.881	.123 [.117, .130]	.056	Large gain over Model 1; acceptable incremental fit, weak absolute fit
3. Higher-order *g*	1,010.45	41	.911	.881	.123 [.117, .130]	.056	Equivalent to Model 2 by construction; not separately testable
4. Bifactor	538.87	33	.954	.923	.099 [.092, .107]	.037	Best fit; GK specific factor collapses to zero variance (Heywood case)

Note. All χ^2 values significant at $p < .001$. CFI = comparative fit index; TLI = Tucker–Lewis index; RMSEA = root mean square error of approximation. Models 2 and 3 are empirically equivalent because the higher-order portion is just-identified with three first-order factors. RMSEA is given with its 90% confidence interval. SRMR = standardized root mean square residual, computed from the observed and model-implied correlation matrices. The conventional cut-offs are not the right yardstick for this design; the calibrated benchmark is reported below. The Interpretation column summarises the reading defended in the text; it is not an additional statistic.

The one-factor model fit poorly (CFI = .786, RMSEA = .185), indicating that a single latent dimension is insufficient to account for the covariance structure among the 11 parcels. The three-factor correlated model fit substantially better (CFI = .911, RMSEA = .123), and the chi-square difference test confirmed a statistically significant improvement ($\Delta\chi^2 = 1{,}370.0$, $\Delta df = 3$, $p < .001$). The higher-order *g* model produced essentially identical fit to the three-factor correlated model, and the reason is structural rather than

coincidental. A second-order factor explains the covariances among the first-order factors; with three first-order factors there are exactly three such covariances and exactly three second-order loadings to estimate, so the second-order layer is just-identified — it reproduces the interfactor correlations perfectly and can neither fit them better nor worse than the correlated model does by estimating them freely. The two models are therefore statistically equivalent here, and the choice between them is theoretical, not empirical: a higher-order model asserts that the domains are related *because* they share a general factor, whereas the correlated model merely records that they are related. With only three domains, the claim that the domains are correlated and the claim that they are correlated because they share a general factor make identical predictions about the data, so no fit statistic can prefer one over the other. The data cannot adjudicate between them; a fourth domain would make the second-order layer overidentified and the comparison meaningful ($\chi^2 = 1{,}010.45$, $df = 41$), which is expected because the higher-order portion of the model is just-identified when there are exactly three first-order factors (Kline, 2016). The bifactor model achieved the best fit (CFI = .954, TLI = .923, RMSEA = .099), although the GK specific factor exhibited a Heywood case (near-zero variance), suggesting that GK parcel variance is almost entirely captured by the general factor.

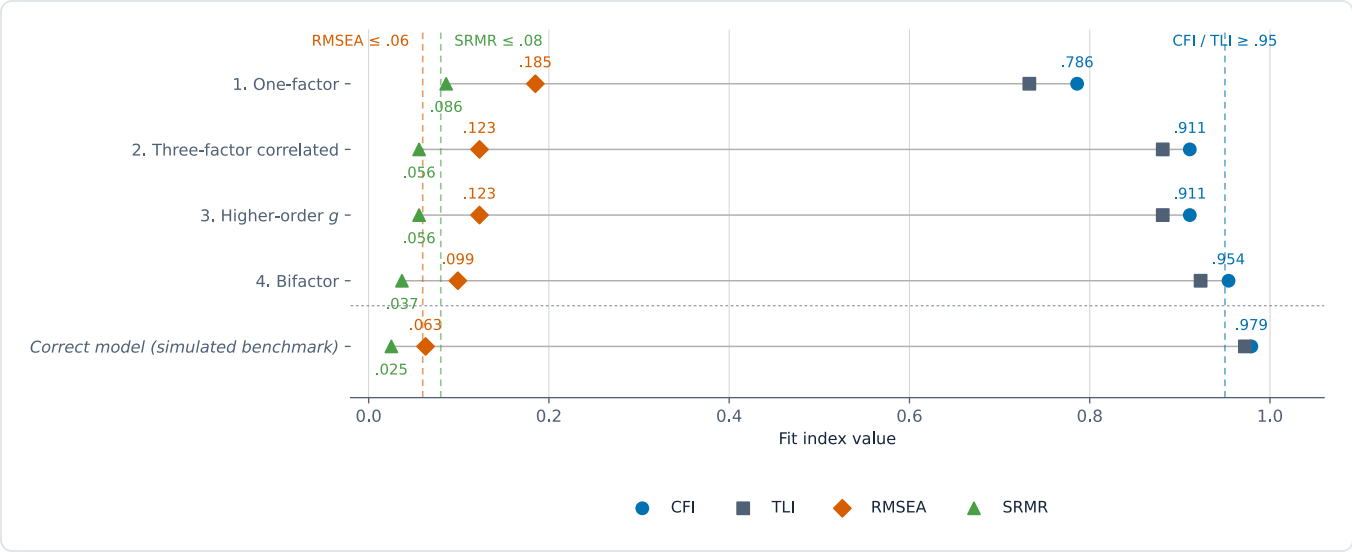

Figure 2. Fit indices for the four competing models (values from Table 2), with the calibrated benchmark beneath the rule: the fit that a model which is *correct by construction* attains in this design, from 200 simulated replications. Dashed lines mark the conventional cut-offs (Hu & Bentler, 1999). The benchmark is the row that matters. A correct model here reaches RMSEA .063, not .00, so the conventional cut-off is unattainable by construction and reading .123 against .06 overstates the case against the model. But every fitted model, the bifactor included, sits worse than the benchmark on all four indices — the three-factor RMSEA and SRMR are each roughly double it — so the misfit is not merely the price of the indicators. Part of it is real.

The RMSEA of the three-factor model, .123, exceeds every conventional threshold, and its confidence interval is narrow enough (.117 to .130) that the value is not in doubt. Two explanations are available. Either the three-factor structure is wrong, or the conventional thresholds are the wrong yardstick for indicators of this kind — parcels built from binary items and analysed by ML, which introduces residual non-normality that inflates chi-square-based indices (Bandalos, 2002). The large sample is not a candidate: it inflates the chi-square statistic but not the RMSEA, which carries its own correction for *N* and degrees of freedom. Papers in this position usually assert the second explanation and move on. It can instead be measured, and the next section measures it.

A Calibrated Benchmark for Absolute Fit

What fit would a *correct* model attain under these conditions? The question is answerable by simulation. Response data were generated from a model that is true by construction — the item-level three-factor solution estimated above, with the real item thresholds and the real loadings — then parcelled exactly as the observed data were parcelled and refitted with the same parcel-level three-factor ML model. The structure is correct in every replication by design, so whatever misfit appears is the misfit that the indicators, the parcelling, and the estimator produce on their own. Two hundred replications at $N = 1{,}553$ give the benchmark.

A correctly specified model, in this design, yields RMSEA = .063 (5th to 95th percentile .053 to .072), CFI = .979, TLI = .972, and SRMR = .025 (.021 to .029). Two conclusions follow, and they cut in opposite directions.

The first is that the conventional cut-offs are miscalibrated here. A model that is *exactly right* cannot reach RMSEA ≤ .06 with these indicators; the standard threshold for good fit is unattainable by construction. Reading .123 against .06, as a referee reasonably would, therefore overstates the case against the model, and the same holds for CFI: a correct model reaches .979, not 1.00.

The second is that this does not exonerate the model. Against the benchmark rather than the convention, the three-factor solution is worse than a correct model on every index — RMSEA .123 against .063, CFI .911 against .979, TLI .881 against .972, SRMR .056 against

.025 — and the gaps are large: the observed RMSEA is roughly double the benchmark, and so is the observed SRMR, while the benchmark's 95th percentile (.072) lies far below the lower bound of the observed confidence interval (.117). The elevated RMSEA is therefore *partly* an artefact of the indicators and *partly* genuine misspecification. The three-factor model is not merely a correct model ill served by its indicators: there is something in the covariance structure that it fails to capture. The next section locates it.

Two limits on this benchmark should be stated. It is derived under a three-factor generating model, so it is the appropriate comparison for the three-factor fit and not, strictly, for the bifactor, whose extra parameters would earn it a benchmark of its own. And it inherits whatever is wrong with the item-level solution used to generate it. It is a calibration, not a proof.

Residual Correlations

No modification indices were consulted and no model was respecified: the four models tested are the four specified in advance, and no residual covariance was freed to improve fit. The residuals themselves, however, are informative about *where* the three-factor model misses. The largest standardized residual correlation is between the two easiest Verbal Analogies parcels (VA_P1 with VA_P2, +.201); the next largest are VA_P1 with GK_P1 (+.133) and MP_P3 with GK_P4 (+.129). The pattern is not random. Because parcels were formed from consecutive blocks of difficulty-ordered items, the first parcel of each subtest contains that subtest's easiest items, and the easiest parcels covary with one another beyond what the content factors imply — a difficulty artefact of the parcelling scheme rather than a substantive fourth factor. This is where the misfit that the benchmark could not explain away appears to live. It is a property of the indicators rather than of the three-factor hypothesis — the covariance is between parcels of like difficulty, not between content domains — and a difficulty-balanced parcelling would be expected to reduce it. That prediction is testable, and it is not tested here.

Standardized Factor Loadings: Three-Factor Correlated Model

Table 3 presents the standardized factor loadings from the three-factor correlated model (Model 2), which represents the most interpretable solution.

Table 3

Standardized Factor Loadings for the Three-Factor Correlated Model

Parcel	Items	VA	MP	GK
VA_P1	10	.513	—	—
VA_P2	10	.694	—	—
VA_P3	11	.836	—	—
VA_P4	10	.740	—	—
MP_P1	10	—	.796	—
MP_P2	11	—	.924	—
MP_P3	11	—	.732	—
GK_P1	14	—	—	.667
GK_P2	14	—	—	.889
GK_P3	14	—	—	.865
GK_P4	14	—	—	.858

Note. All loadings significant at $p < .001$. Dashes indicate fixed-to-zero loadings.

All standardized loadings were substantial, ranging from .513 (VA_P1) to .924 (MP_P2). Within the VA factor, loadings increased monotonically from P1 (.513) to P3 (.836), reflecting higher factor saturation in the more difficult item parcels. The MP factor showed the highest average loading (.817), with MP_P2 achieving the strongest single loading (.924) in the model. The GK factor loadings were uniformly high (.667 to .889), with the easier parcel (GK_P1, $M = 11.35$ of 14) showing the lowest loading.

Interfactor Correlations

Table 4 presents the interfactor correlation matrix from the three-factor correlated model.

Table 4

Interfactor Correlations From the Three-Factor Correlated Model

Factor	VA	MP	GK
VA	—		
MP	.694	—	
GK	.890	.609	—

Note. All correlations significant at $p < .001$.

The VA–GK correlation (.890) was the highest, reflecting that both subtests heavily sample verbal–educational knowledge consistent with the *Gc* construct. The VA–MP (.694) and MP–GK (.609) correlations were lower, indicating that mathematical reasoning shares substantial variance with verbal and knowledge abilities but retains a larger component of unique variance. The pattern of interfactor correlations is consistent with a dominant general factor that explains most covariance, with mathematical ability representing a partially distinct facet.

Schmid–Leiman Decomposition and Omega Indices

The Schmid–Leiman (S-L) transformation was applied to the three-factor correlated solution to partition each parcel's variance into general and specific components. Table 5 presents the S-L general factor loadings alongside the specific factor residual loadings, and Figure 3 shows the estimated solution: the three-factor model with its loadings in panel (a), and the general–specific partition of each parcel in panel (b).

Table 5
Schmid–Leiman General and Specific Factor Loadings

Parcel	*g*	sVA	sMP	sGK
VA_P1	.490	.151	—	—
VA_P2	.663	.204	—	—
VA_P3	.799	.246	—	—
VA_P4	.707	.218	—	—

Parcel	*g*	sVA	sMP	sGK
MP_P1	.665	—	.437	—
MP_P2	.772	—	.508	—
MP_P3	.612	—	.402	—
GK_P1	.617	—	—	.252
GK_P2	.823	—	—	.336
GK_P3	.801	—	—	.327
GK_P4	.794	—	—	.325

Note. g = general factor loading; sVA, sMP, sGK = specific factor residual loadings after extraction of *g*. Loadings derived from Schmid–Leiman orthogonalization of the three-factor correlated model.

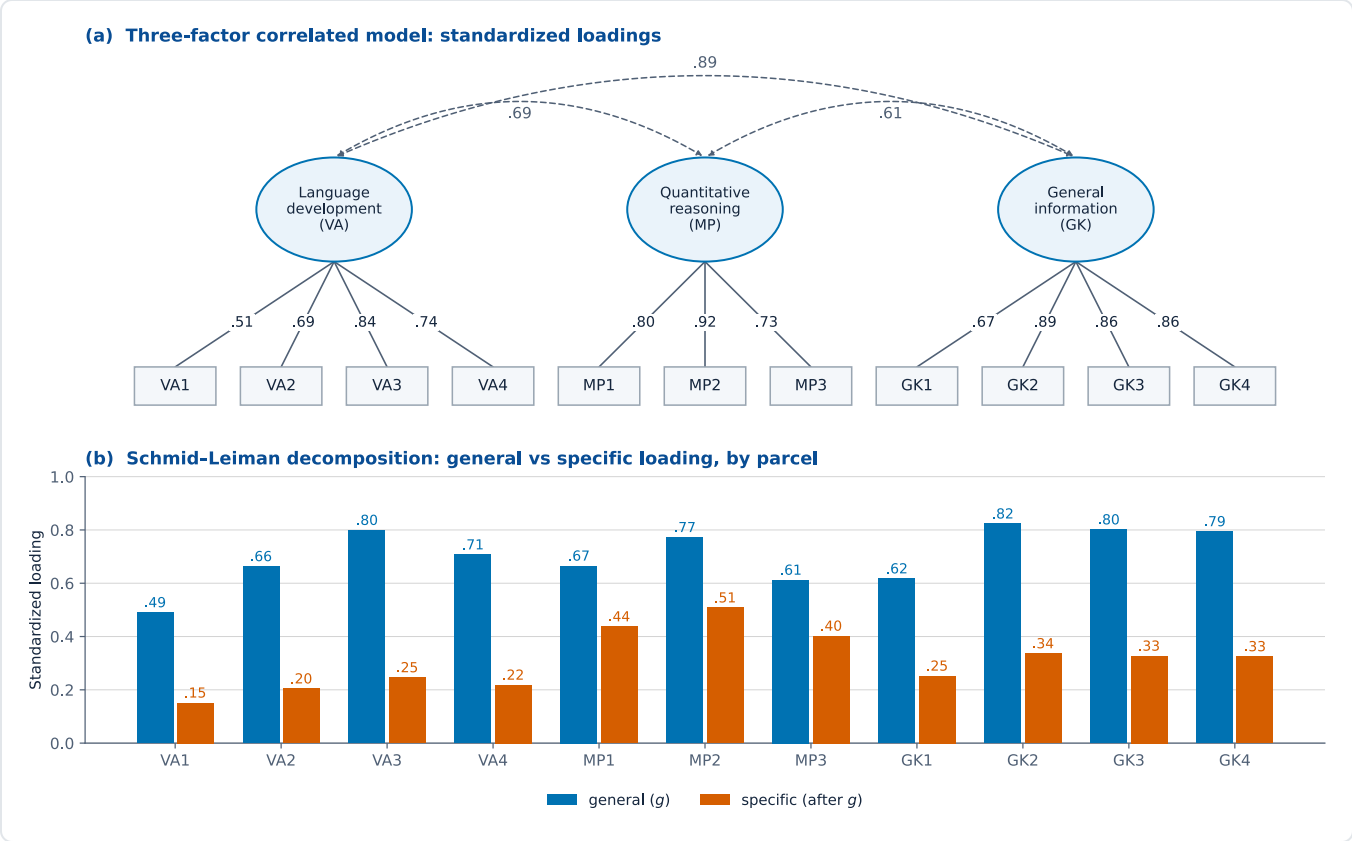

Figure 3. The estimated solution. (a) The three-factor correlated model with standardized loadings on each path and interfactor correlations on the dashed arrows (Tables 3 and 4). The two verbal–educational facets of crystallized ability — language development (VA) and general verbal information (GK) — correlate .89, while quantitative reasoning (MP) stands apart at .69 and .61. (b) The Schmid–Leiman decomposition of that solution (Table 5): for each parcel, the loading on the general factor beside the loading on its own specific factor, once *g* has been extracted. The bars are aligned under the parcels of panel (a). The general loading exceeds the specific loading for every one of the 11 parcels, which is

what ω_h = .879 and ECV = .826 record numerically; the specific factors retain most of what little is left in Mathematical Problems and least in Verbal Analogies.

General factor loadings ranged from .490 (VA_P1) to .823 (GK_P2), with a median of .707. Specific factor loadings were substantially smaller: sVA loadings ranged from .151 to .246, sGK from .252 to .336, and sMP from .402 to .508. The MP parcels showed the largest specific factor loadings relative to their *g* loadings, confirming that mathematical reasoning retains the most unique variance after controlling for the general factor.

The omega indices derived from the S-L transformation are presented in Table 6.

Table 6

Omega Hierarchical, Omega Subscale, and ECV Indices

Index	Value
ω_h (total score)	.879
ω_t (total score)	.937
ω_h / ω_t (share of *reliable* variance due to *g*)	.937
ECV (*g*)	.826
ω_s (VA)	.069
ω_s (MP)	.260
ω_s (GK)	.128
% *g* in VA subscale	91.4%
% *g* in MP subscale	69.8%
% *g* in GK subscale	85.7%

Note. ω_h = omega hierarchical (proportion of *total* score variance attributable to *g*); ω_t = omega total (proportion of total score variance that is reliable); ω_h/ω_t = the proportion of the *reliable* variance attributable to *g*; ECV = explained common variance (proportion of *common* variance due to *g*); ω_s = omega subscale (proportion of subscale variance attributable to the specific factor after controlling for *g*); % *g* = proportion of reliable subscale variance due to *g*. Each factor contributes the *square of the sum* of its loadings to the composite variance, not the sum of its squared loadings.

The omega hierarchical of .879 indicates that 87.9% of the variance in total scores is attributable to the general factor. Since $\omega_t = .937$ of that variance is reliable at all, the general factor claims 93.7% of the *reliable* variance (ω_h/ω_t) — the quantity of interest for score interpretation, and the one it is easy to conflate with ω_h itself. The ECV of .826 confirms that 82.6% of the common variance is due to *g*, well above the .70 threshold typically used to support essentially unidimensional interpretation (Rodriguez et al., 2016). Omega subscale values reveal that subtest scores capture minimal unique reliable variance beyond *g*: VA ($\omega_s = .069$) and GK ($\omega_s = .128$) are almost entirely determined by the general factor, while MP ($\omega_s = .260$) retains modest but non-trivial unique variance. The percentage of reliable variance attributable to *g* in each subscale further illustrates this pattern: 91.4% for VA, 85.7% for GK, and 69.8% for MP.

Item-Level Replication With a Categorical Estimator

The parcel-level ML analysis rests on two concessions — parcelling and a continuous-data estimator applied to indicators built from binary items — and the natural question is whether either concession is doing the work. To answer it, the four models were refitted at the item level to the tetrachoric correlation matrix of all 129 binary items, estimated by ULS.

Two properties of that matrix are worth stating, because they bear on why an item-level analysis was not the first choice. Twenty-six of the 8,256 item pairs produced an empty cell in their 2×2 contingency table — a combination of responses that no examinee in 1,553 gave — which drives the corresponding tetrachoric correlation to the boundary. All 26 involve one of six extreme items (three passed by fewer than 5% of examinees, three by more than 95%). The matrix as estimated is consequently not positive definite (19 negative eigenvalues), and was smoothed to the nearest positive-definite correlation matrix before fitting (Higham, 2002). Item-level categorical estimation on this pool is therefore possible but not comfortable, which is itself part of the case for parcelling.

With that caveat, the replication is reassuring. The ordering of the models by SRMR is exactly the ordering found at the parcel level: one-factor .092, three-factor correlated .076, higher-order *g* .076 (again indistinguishable from the correlated model, as it must

be), bifactor .068. The interfactor correlations barely move — VA–MP = .691 against .694 at the parcel level, VA–GK = .878 against .890, MP–GK = .630 against .609 — so the asymmetric pattern on which the substantive interpretation rests, with the two verbal–educational domains close together and the quantitative domain standing apart, survives the change of estimator intact. Most importantly, the explained common variance is .822 at the item level against .826 at the parcel level. ECV is the right quantity for this comparison because it is a ratio of common variances and so does not depend on the number of indicators; ω_h, by contrast, rises mechanically with test length (its numerator is quadratic in the number of indicators while the residual sum is linear), and item-level and parcel-level values of it are therefore not comparable and are not compared here.

Nor does any of this rest on the smoothing. Refitting on the 123 items that remain when the six extreme items are removed — a matrix that needs far less repair, its cells moving by .006 on average rather than .016 — changes nothing that matters: SRMR .088, .070, .070, and .062 across the four models, interfactor correlations of .684, .869, and .627, and an ECV of .819. The three estimates of the explained common variance, from three different treatments of the data, are .826, .822, and .819.

The conclusions of this paper do not, then, depend on the parcelling, on the choice of ML, or on the repair of the tetrachoric matrix. They are properties of the data.

Discussion

RQ1. Confirmatory factor analysis rejects the unidimensional model in favor of a three-factor structure (VA, MP, GK) corresponding to the three content domains. **RQ2.** A dominant general factor accounts for 93.7% of the reliable variance ($\omega_h = .879$, $\omega_t = .937$, ECV = .826); the VA and GK subtests are largely subsumed by the general factor, while MP retains modest unique variance ($\omega_s = .260$).

Three-Factor Structure With a Dominant General Factor

The present study provides the first confirmatory factor analytic examination of the JCCES latent structure. The results clearly reject the unidimensional model in favor of a

three-factor structure corresponding to the three content domains (VA, MP, GK), while simultaneously revealing that a dominant general factor accounts for the vast majority of reliable variance. This pattern—a statistically necessary multifactor structure coexisting with a psychologically dominant general factor—is characteristic of well-constructed cognitive test batteries (Carroll, 1993; Johnson et al., 2004) and is precisely what the CHC theory predicts for measures of broad *Gc*.

The chi-square difference test between the unidimensional and three-factor models ($\Delta\chi^2 = 1{,}370.0$, $\Delta df = 3$, $p < .001$) is corroborated by the fit-index changes, which do not depend on sample size in the same way: $\Delta\text{CFI} = .125$ and $\Delta\text{RMSEA} = -.062$, both an order of magnitude beyond the .01 that conventionally marks a non-trivial improvement (Cheung & Rensvold, 2002). The evidence therefore indicates that the three content domains are not interchangeable indicators of a single factor. Nevertheless, the high interfactor correlations (mean $r = .731$) and the dominance of the general factor in the Schmid–Leiman decomposition (ECV = .827) indicate that this multidimensionality is nested within a strong general factor. The practical consequence is that total scores are well-supported by a single construct, while subtest differences carry limited interpretive weight.

The interfactor correlations provide insight into the internal organization of crystallized intelligence as measured by the JCCES. The very high VA–GK correlation (.890) indicates that verbal analogical reasoning and factual knowledge are nearly indistinguishable at the latent level, consistent with both abilities reflecting accumulated verbal–educational experience. In contrast, MP's lower correlations with VA (.694) and GK (.609) suggest that quantitative reasoning, while strongly related to the general crystallized factor, involves partially distinct cognitive processes—plausibly quantitative reasoning (*Gq*) in the CHC framework (McGrew, 2009). This asymmetric correlation pattern, with verbal–knowledge subtests clustering together and the quantitative subtest standing apart, is broadly consistent with findings from major individually administered intelligence batteries such as the WAIS-IV Verbal Comprehension Index and its subtests.

Implications for Score Interpretation

The omega hierarchical of .879 has direct implications for JCCES score interpretation. Following the framework of Rodriguez et al. (2016), an ω_h above .80 supports interpreting total scores as primarily reflecting a single construct. In the case of the JCCES, total scores (CAI) can be interpreted with high confidence as a measure of general crystallized ability. The low ω_s values for VA (.069) and GK (.128) indicate that these subtest scores carry almost no unique reliable information beyond what is already captured by the total score. Clinicians and researchers should be cautious about interpreting VA–GK subtest differences, as such differences are overwhelmingly attributable to measurement error rather than true ability differences.

The MP subtest stands apart with an ω_s of .260, indicating that approximately 26% of the reliable variance in MP scores reflects a quantitative ability component that is genuinely distinct from the general crystallized factor. This level of unique variance, while modest, may be sufficient to support interpretive inferences when MP scores diverge substantially from the overall pattern. The finding is consistent with the position of quantitative reasoning as a partially separate narrow ability within the CHC taxonomy (McGrew, 2009).

In practical terms, these results recommend a hierarchical interpretive strategy for the JCCES: (a) interpret the total score (CAI) as the primary measure of general crystallized ability; (b) examine the MP subtest score for potential discrepancies from the overall profile, as such discrepancies may reflect genuine differences in quantitative versus verbal–educational abilities; and (c) avoid interpreting VA–GK differences clinically, as the very low ω_s values for these subtests indicate that discrepancies between them are dominated by measurement error.

Comparison With Previous Structural Analyses of the JCCES

A previous multidimensional scaling (MDS) analysis examined how the three JCCES subtests and the four self-reported ACT sections organize in a low-dimensional similarity space, in a sample restricted to examinees whose ACT was plausibly the post-1989 Enhanced ACT ($n = 57$; Jouve, 2016). The two-dimensional solution (Stress-1 = .073, RSQ = .960) showed two organizations, each supported by an exact permutation test

independent of axis orientation and directionally stable in all 1,000 bootstrap resamples: a content contrast running from verbal and knowledge tasks to quantitative and scientific-reasoning tasks, and a slightly larger separation of the JCCES subtests from the ACT sections — an instrument-level organization that, read descriptively as a multitrait–multimethod matrix, reflects same-instrument correlations exceeding content-matched cross-instrument convergence. The present CFA results are consistent with the content contrast: the high VA–GK interfactor correlation (.890) corresponds to the proximity of the two verbal–knowledge subtests at one end of that contrast, while MP's lower correlations with VA (.694) and GK (.609) mirror its position toward the quantitative pole. The CFA complements the MDS account by quantifying the within-battery structure through standardized loadings, interfactor correlations, and omega indices, on a substantially larger single-battery sample ($N = 1{,}553$ vs. $n = 57$).

The Bifactor Model and Practical Considerations

The bifactor model (Model 4) achieved the best absolute fit (CFI = .954), but the GK specific factor exhibited a Heywood case (near-zero variance), indicating that GK's covariance structure is entirely explained by the general factor. This pattern—a well-fitting bifactor model with one or more collapsed specific factors—is common when content domains are highly correlated and has been documented in analyses of other intelligence batteries (Beaujean, 2014; Reise, 2012). A Heywood case of this kind should not be read as numerical instability or a software failure. The model asks the data to support a GK-specific factor *over and above* the general factor; the estimate of that factor's variance goes to zero because there is no such residual covariance to explain — GK's items covary with one another no more than *g* already implies. The boundary estimate is thus the model reporting that it has been given a parameter it does not need: the bifactor specification is overparameterized for these data, and the collapse is informative rather than pathological. It says the same thing as the omega and ECV results by another route, and it is why the three-factor correlated model, not the bifactor model, is the one interpreted here despite the latter's better absolute fit.

Limitations

Two design decisions bound what this analysis can claim, and both are concessions rather than ideals. The first is parcelling. Aggregating 129 binary items into 11 parcel sums buys normality and a tractable model, but it buys them by averaging away whatever structure exists *within* a subtest. A parcel is unidimensional by construction; if the Verbal Analogies items in fact tap two distinguishable abilities, the parcels will hide that, and this study would not detect it. Nothing reported here should therefore be read as evidence that the JCCES subtests are internally unidimensional — only that, at the domain level, three factors are needed and one is not. The second is the estimator. Binary indicators call for least squares on tetrachoric correlations rather than ML (Muthén & Muthén, 2017), and ML on parcel sums is a defensible approximation at the domain level, not the correct estimator for the underlying data. That objection is no longer merely conceded: the item-level ULS replication reported above answers it directly, and the substantive conclusions survive. What the replication does not supply is a mean- and variance-adjusted test statistic, so it cannot furnish the absolute-fit verdict that a full WLSMV analysis in software built for it would. The absolute fit of these models remains the part of the evidence the reader should trust least; the comparative evidence, which now holds under two estimators at two levels of aggregation, is the part the design supports.

The sample carries a third limitation. The online convenience sample may not be representative of the general population: self-selected examinees who seek out cognitive assessment may differ from normative samples in motivation, education, or ability distribution, and the structure recovered here is the structure of that population.

Finally, the RMSEA values (.123 for the three-factor model, .099 for the bifactor) exceed conventional cutoffs for good fit. As set out above, this is not an artefact of the large sample — RMSEA carries its own correction for N and degrees of freedom — and the simulated benchmark shows that roughly half of it is the price of the indicators and the parcelling, the remainder being genuine misspecification — located, so far as the residuals reveal, in the covariance among the easiest parcels rather than among the content domains.

Future Directions

Several avenues for future research are suggested by the present findings. First, the item-level replication reported above should be carried further in software that supplies the mean- and variance-adjusted statistics of a full WLSMV analysis, which would furnish the absolute-fit verdict this paper cannot. The same analysis would allow within-subtest dimensionality to be examined directly, which parcelling forecloses. A second and cheaper check follows from the residual correlations: re-parcelling on a difficulty-balanced rather than a consecutive-block scheme should attenuate the covariance among the easiest parcels, and would show how much of the elevated RMSEA is an artefact of the indicators. Second, invariance testing across demographic subgroups (e.g., gender, education level, age) would establish whether the three-factor structure holds equivalently across populations, a prerequisite for fair comparison of subtest profiles. Third, the convergent and discriminant validity of the JCCES factor structure could be examined by fitting joint CFA models incorporating external measures of fluid intelligence (*Gf*), quantitative reasoning (*Gq*), and crystallized ability (*Gc*) from established batteries, thereby situating the JCCES within the broader CHC nomological network. Fourth, cross-validation with independent samples from diverse cultural and linguistic backgrounds would strengthen the generalizability of the three-factor structure reported here.

Conclusion

The JCCES exhibits a three-factor structure corresponding to its Verbal Analogies, Mathematical Problems, and General Knowledge subtests, with a dominant general factor that accounts for 93.7% of the reliable variance in total scores ($\omega_h = .879$, $\omega_t = .937$, ECV = .826). The VA and GK subtests are largely subsumed by the general factor, while the MP subtest retains modest unique variance ($\omega_s = .260$) consistent with partially distinct quantitative reasoning abilities. These results support interpreting JCCES total scores (CAI) as a reliable measure of general crystallized intelligence, with the caveat that VA–GK subtest differences should not be overinterpreted. The MP subtest score may provide incremental information about quantitative ability when it diverges substantially from the overall profile.

Figures & tables

Figure 1. *The Four Competing Measurement Models, as Specified*

Figure 2. *Fit Indices for the Four Competing Models*

Figure 3. *The Estimated Solution: Standardized Loadings and the Schmid–Leiman Decomposition*

Table 1. *Descriptive Statistics for JCCES Subtests and Total Score (N = 1,553)*

Table 2. *Fit Statistics for Competing CFA Models (N = 1,553)*

Table 3. *Standardized Factor Loadings for the Three-Factor Correlated Model*

Table 4. *Interfactor Correlations From the Three-Factor Correlated Model*

Table 5. *Schmid–Leiman General and Specific Factor Loadings*

Table 6. *Omega Hierarchical, Omega Subscale, and ECV Indices*

References

Bandalos, D. L. (2002). The effects of item parceling on goodness-of-fit and parameter estimate bias in structural equation modeling. *Structural Equation Modeling, 9*(1), 78–102. https://doi.org/10.1207/S15328007SEM0901_5

Beaujean, A. A. (2014). *Latent variable modeling using R: A step-by-step guide*. Routledge.

Bollen, K. A. (1989). *Structural equations with latent variables*. Wiley. https://doi.org/10.1002/9781118619179

Carroll, J. B. (1993). *Human cognitive abilities: A survey of factor-analytic studies*. Cambridge University Press. https://doi.org/10.1017/CBO9780511571312

Cattell, R. B. (1963). Theory of fluid and crystallized intelligence: A critical experiment. *Journal of Educational Psychology, 54*(1), 1–22. https://doi.org/10.1037/h0046743

Cheung, G. W., & Rensvold, R. B. (2002). Evaluating goodness-of-fit indexes for testing measurement invariance. *Structural Equation Modeling*, *9*(2), 233–255. https://doi.org/10.1207/s15328007sem0902_5

Forero, C. G., Maydeu-Olivares, A., & Gallardo-Pujol, D. (2009). Factor analysis with ordinal indicators: A Monte Carlo study comparing DWLS and ULS estimation. *Structural Equation Modeling, 16*(4), 625–641. https://doi.org/10.1080/10705510903203573

Higham, N. J. (2002). Computing the nearest correlation matrix — A problem from finance. *IMA Journal of Numerical Analysis, 22*(3), 329–343. https://doi.org/10.1093/imanum/22.3.329

Hu, L., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling, 6*(1), 1–55. https://doi.org/10.1080/10705519909540118

Igolkina, A. A., & Meshcheryakov, G. (2020). semopy: A Python package for structural equation modeling. *Structural Equation Modeling, 27*(6), 952–963. https://doi.org/10.1080/10705511.2019.1704289

Johnson, W., Bouchard, T. J., Krueger, R. F., McGue, M., & Gottesman, I. I. (2004). Just one *g*: Consistent results from three test batteries. *Intelligence, 32*(1), 95–107. https://doi.org/10.1016/S0160-2896(03)00062-X

Jöreskog, K. G. (1969). A general approach to confirmatory maximum likelihood factor analysis. *Psychometrika, 34*(2), 183–202. https://doi.org/10.1007/BF02289343

Jouve, X. (2016). Content and instrument organization of JCCES and ACT subtests: A multidimensional scaling analysis. *Cogn-IQ Research Papers*. https://pubscience.org/ps-1mADL-b45fe5-Sjy6

Jouve, X. (2023). Evaluating the Jouve Cerebrals Crystallized Educational Scale (JCCES): Reliability, internal consistency, and alternative answer recognition. *Cogn-IQ Research Papers*. https://pubscience.org/ps-1mSR3-32426c-iYHT

Jouve, X., Wai, J., Watterson, J., Woodley, M., Faverio, M. C., Lindgren, K., Whitley, J., & Freitas, N. (2023). *Jouve-Cerebrals Crystallized-Educational Scale (JCCES) – Revised*. Cogn-IQ Cognitive Assessments. https://pubscience.org/ps-1mVDO-e10f3a-Tc1I

Kline, R. B. (2016). *Principles and practice of structural equation modeling* (4th ed.). Guilford Press.

Little, T. D., Cunningham, W. A., Shahar, G., & Widaman, K. F. (2002). To parcel or not to parcel: Exploring the question, weighing the merits. *Structural Equation Modeling, 9*(2), 151–173. https://doi.org/10.1207/S15328007SEM0902_1

Matsunaga, M. (2008). Item parceling in structural equation modeling: A primer. *Communication Methods and Measures, 2*(4), 260–293. https://doi.org/10.1080/19312450802458935

McGrew, K. S. (2009). CHC theory and the human cognitive abilities project: Standing on the shoulders of the giants of psychometric intelligence research. *Intelligence, 37*(1), 1–10. https://doi.org/10.1016/j.intell.2008.08.004

Muthén, L. K., & Muthén, B. O. (2017). *Mplus user's guide* (8th ed.). Muthén & Muthén.

Reise, S. P. (2012). The rediscovery of bifactor measurement models. *Multivariate Behavioral Research*, *47*(5), 667–696. https://doi.org/10.1080/00273171.2012.715555

Rodriguez, A., Reise, S. P., & Haviland, M. G. (2016). Evaluating bifactor models: Calculating and interpreting statistical indices. *Psychological Methods, 21*(2), 137–150. https://doi.org/10.1037/met0000045

Schmid, J., & Leiman, J. M. (1957). The development of hierarchical factor solutions. *Psychometrika, 22*(1), 53–61. https://doi.org/10.1007/BF02289209

Zinbarg, R. E., Revelle, W., Yovel, I., & Li, W. (2005). Cronbach's α, Revelle's β, and McDonald's ω_H: Their relations with each other and two alternative conceptualizations of reliability. *Psychometrika, 70*(1), 123–133. https://doi.org/10.1007/s11336-003-0974-7

CITE THIS ARTICLE

Jouve, X. (2026). Confirmatory factor analysis of the Jouve Cerebrals Crystallized Educational Scale: Evidence for a three-factor structure with a dominant general factor. *Cogn-IQ Research Papers.* https://pubscience.org/ps-1n0I3-799113-Lpri

Export: BibTeX · RIS · EndNote · APA · Plain text

OUTLINE

Abstract

Introduction

CHC & Crystallized Intelligence

CFA & the LISREL Tradition

The Present Study

Method

- Factor Analysis — Statistical techniques for identifying latent constructs in psychometric data
- CHC Theory — The Cattell-Horn-Carroll model of cognitive abilities
- Crystallized Intelligence — Accumulated knowledge and verbal skills acquired through education and experience